UC Irvine
UC Irvine Previously Published Works

Title

Evidence of Nonrandom Patterns of Functional Chromosome Organization in Danaus plexippus.

Permalink

https://escholarship.org/uc/item/62r5k09f

Authors

Kimura, Ashlyn
Go, Alwyn C
Markow, Therese
et al.

Publication Date

2024-03-01

DOI

10.1093/gbe/evae054

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial License, available at https://creativecommons.org/licenses/by-nc/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Article

Evidence of Nonrandom Patterns of Functional Chromosome Organization in *Danaus plexippus*

Ashlyn Kimura [1], Alwyn C. Go [2], Therese Markow [3,4], José M. Ranz [1]*

[1] Department of Ecology and Evolutionary Biology, University of California Irvine, Irvine CA 92647, USA

[2] Department of Biology, University of Winnipeg, Winnipeg, MB R3B 2E9, Canada

[3] Unidad de Genómica Avanzada (Langebio), CINVESTAV, Irapuato GTO 36824, México

[4] Section of Cell and Developmental Biology, Division of Biological Sciences, University of California San Diego, La Jolla CA 92093, USA

* Corresponding author (jranz@uci.edu)

Running title: functional gene organization in the monarch chromosomes

ABSTRACT

Our understanding on the interplay between gene functionality and gene arrangement at different chromosome scales relies on a few Diptera and the honeybee, species with quality-reference genome assemblies, accurate gene annotations, and abundant transcriptome data. Using recently generated 'omics resources in the monarch butterfly *D. plexippus*, a species with many more and smaller chromosomes relative to *Drosophila* species and the honeybee, we examined the organization of genes preferentially expressed at broadly defined developmental stages (larva, pupa, adult males, and adult females) at both fine and whole-chromosome scales. We found that developmental stage-regulated genes do not form more clusters, but do form larger clusters, than expected by chance, a pattern consistent across the gene categories examined. Notably, out of the thirty chromosomes in the monarch genome, twelve of them, plus the fraction of the chromosome *Z* that corresponds to the ancestral *Z* in other Lepidoptera, were found enriched for developmental stage-regulated genes. These two levels of nonrandom gene organization are not independent as enriched chromosomes for developmental stage-regulated genes tend to harbor disproportionately large clusters of these genes. Further, although paralogous genes were overrepresented in gene clusters, their presence is not enough to explain two-thirds of the documented cases of whole-chromosome enrichment. The composition of the largest clusters often included paralogs from more than one multigene family as well as unrelated single-copy genes. Our results reveal intriguing patterns at the whole-chromosome scale in *D. plexippus* while shedding light on the interplay between gene expression and chromosome organization beyond Diptera and Hymenoptera.

Keywords: chromosome organization, gene clustering, expression profiles, sex-biased expressed genes, Lepidoptera, insects

SIGNIFICANCE STATEMENT

In eukaryotes, chromosome location and gene cluster formation are nonrandom properties often influenced by expression attributes. In insects, this topic has been examined in closely related Dipteran species and the honeybee. In the Lepidopteran *D. plexippus,* a species with 31 chromosomal elements of varying size, we analyzed how genes with different expression trends across the species' life cycle are organized at fine- and whole-chromosome scales. We found robust patterns of nonrandom gene organization at both scales, notably with a large fraction of monarch chromosomes showing enrichment for genes with the same expression trend. Together, our findings highlight how different gene function, assessed here as developmental stage-regulated expression, is intertwined at different scales with the chromosome organization in *D. plexippus*.

INTRODUCTION

In eukaryotes, gene location across the genome is not entirely random (Hurst, et al. 2004). Genes with similar expression profiles often colocalize in the same genomic neighborhood, a feature observed in model organisms and humans (Boutanaev, et al. 2002; Lercher, et al. 2002; Semon and Duret 2006; Williams and Bowles 2004). This coexpression results primarily from a variety of mechanisms, including bidirectional promoters, shared local chromatin states or regulatory sequences, and exposure to promiscuous *cis*-regulatory elements or to the same set of trans-acting factors (Kustatscher, et al. 2017; Zinani, et al. 2022). Although clustering of coexpressed genes has been found for functionally related genes such as those participating in the same pathway (Lee and Sonnhammer 2003), it often involves non-functionally related genes (Kustatscher, et al. 2017; Weber and Hurst 2011; Williams and Bowles 2004). Expression similarity between neighboring genes decreases with physical distance (Quintero-Cadena and Sternberg 2016) although long-range coregulation in *cis* is well documented (Ghavi-Helm, et al. 2014; Kustatscher, et al. 2017). The basic notion that expression similarity is adaptive, thus explaining a sizable fraction of the clustering among coexpressed genes, has been increasingly challenged. This has been the case when how structural variants impact the integrity of coexpression clusters is considered (Weber and Hurst 2011), when coexpression clusters are disrupted with genome engineering tools (Meadows, et al. 2010), or when protein levels –and not only mRNA-levels– are analyzed (Kustatscher, et al. 2017). Nevertheless, clusters of coexpressed genes represent a common feature to the genome organization in eukaryotes, having important implications for gene regulation, being perhaps beneficial for other reasons such as expression noise reduction (Kustatscher, et al. 2017; Zinani, et al. 2022).

Despite the large share that insects represent relative to all eukaryotic diversity, the interplay between chromosomal gene organization and gene functionality has been primarily examined in *Drosophila melanogaster* and some of its close relatives (Boutanaev, et al. 2002; Mezey, et al.

2008; Weber and Hurst 2011), and in only one non-Dipteran species, *Apis mellifera* (Duncan, et al. 2020). The order Lepidoptera (butterflies and moths) accounts for 11.34% of all animal species (Banki, et al. 2022). In lepidopterans, the haploid karyotype mode is 31 and thought to reflect the ancestral chromosomal complement in this order (Robinson 1971), being substantially higher than that of *Drosophila* species (n=4-6) and *A. mellifera* (n=16). Lepidopterans possess holocentric chromosomes sometimes coupled with inverted meiosis (Lukhtanov, et al. 2018; Mandrioli and Manicardi 2020), and exhibit achiasmatic meiosis in the females (de Vos, et al. 2020; Marec 1996). With few exceptions (Hill, et al. 2019; Mackintosh, et al. 2022), the gene content of the Lepidopteran chromosomes is well conserved even among distantly related lineages (Ahola, et al. 2014; Heliconius Genome, et al. 2012; Hill, et al. 2019; Yasukochi, et al. 2006). Although Lepidoptera genomics has developed substantially in the last two decades (Ellis, et al. 2021), the functional aspects of gene organization in the genome of this species order remain elusive. Notably, in *Pieris napi*, one of the few species in which synteny conservation does not hold (Hill, et al. 2019), it was documented that genome regions with conserved gene order between this species and *Bombyx mori* were enriched for genes with particular functional gene annotation terms, pointing to some sort of functional constraint shaping the evolution of the chromosomal gene organization between the two species. Whether additional patterns of nonrandom gene organization exist in connection with the transcriptional program in Lepidotera in which synteny conservation holds is unknown.

Recently, a highly contiguous genome assembly (DpMex_v1), an enhanced gene annotation (OGS1_DpMex), and a transcriptome atlas have been generated in the iconic species *Danaus plexippus* (Ranz, et al. 2021b). Leveraging these resources, we aimed at understanding the interplay between chromosomal gene organization and gene functionality during the life cycle of this species. The karyotype of *D. plexippus* consists of 30 chromosomes, with the *Z* chromosome being the result of a fusion between the ancestral heterochromosome *Z* to the

lepidopterans and an ancestral autosome (Mongue, et al. 2017). Although generally small, the chromosomes of the monarch and other Lepidoptera differ substantially not only in their length but also in gene number and density (Ranz, et al. 2021b). Here we address whether gene functionality, assessed as developmental stage-regulated expression, is randomly distributed both between and within *D. plexippus* chromosomes, finding conspicuous evidence that it is not the case. Our findings add to those previously reported in Diptera and Hymenoptera, pointing to a common feature about how a fraction of genes is nonrandomly organized in the genome of holometabolous insects.

RESULTS

We took advantage of a recently constructed RNA-Seq-based transcriptome atlas in *D. plexippus* (Ranz, et al. 2021b). This atlas included four types of RNA samples obtained from individuals belonging to different larval and pupal stages, and from anatomical parts of adult male and female individuals, thus defining four broad developmental stages (larva, pupa, males, and females). These RNA samples, although ribodepleted, were not polyA enriched, hence more likely including transcripts of lncRNA genes, which have been shown to be relevant for gene regulation and phenotypic change (Bonasio and Shiekhattar 2014; Wen, et al. 2016; Zhu, et al. 2021). Out of 14,685 genes considered, 2732 (2605 coding and 127 lncRNA) were found to be preferentially expressed (5% False-Discovery Rate, FDR, and a fold-change≥2) at one of the broadly defined developmental stages considered: 1174 during larva stage; 835 during pupa stage; 582 in adult males; and 141 in adult females (Ranz, et al. 2021b)(Methods). We tested for the nonrandom distribution of these developmental stage-regulated genes within and across the chromosomes of *D. plexippus* (fig. 1).

Developmental stage-regulated genes form larger clusters than expected by chance

We characterized clustering properties of developmental stage-regulated genes at a fine-chromosome scale and three threshold distances between each two neighboring genes part of the same cluster. This distance was measured as the number of intervening genes: ≤1 gene; ≤5 genes; and ≤10 genes. Examining clustering properties at three distances accounts for expression similarity not only resulting from genes belonging to a similar chromatin domain or exposed to a common regulatory environment (Kustatscher, et al. 2017; Szabo, et al. 2019), but also from long-range coregulation (Ghavi-Helm, et al. 2014; Kustatscher, et al. 2017).

We performed Monte Carlo simulations in which the gene order was shuffled within each chromosome ($n=1\times10^5$; nominal adjusted *p*-value=0.05; identical settings were applied to all subsequent sets of simulations) to determine the extent to which the observed patterns of gene clustering at a fine-chromosome scale can be found by chance alone (supplementary table 1, Supplementary Material online). Nearly half (1353/2732) of all developmental stage-regulated genes formed clusters at a distance of ≤1 intervening gene ($P_{adj}<1.25\times10^{-4}$; average expected=654). This percentage increases to 77.4% (2144/2732) when the threshold distance in number of intervening genes increases up to 10 genes ($P_{adj}<1.25\times10^{-4}$; average expected=1989). When comparing different expression biases, genes preferentially expressed in larval and pupa stages cluster significantly more often than genes preferentially expressed in adult males and females (4-sample test for equality of proportions, $P_{adj}<1x10^{-3}$ for each threshold distance; fig. 2a; supplementary table 2, Supplementary Material online). Across increasingly higher threshold distances, the proportion of genes with preferential expression in larva, pupa, and adult males significantly increases, a pattern not observed for female preferentially expressed genes (3-sample test for equality of proportions, $P_{adj}<1x10^{-3}$ for each expression bias but for female preferentially expressed genes; fig. 2a and supplementary table 3, Supplementary Material online).

This tendency of developmental stage-regulated genes to cluster within chromosomes is also reflected for some expression biases and threshold distances in the form of cluster sizes being larger than the maximum expected by chance (supplementary table 1, Supplementary Material online) but not in the form of a disproportionally high number of clusters. Only in the case of the threshold distance ≤1, the number of clusters is higher than expected by chance (observed=481 vs average expected=300; $P_{adj}<3\times10^{-4}$; at ≤5: 579 vs average expected=561, $P_{adj}=0.16$; at ≤10: 574 vs average expected=608, $P_{adj}=0.999$; fig. 2b; supplementary table 1, Supplementary Material online). This absence of significant increase in number of clusters at higher threshold distances is consistent across expression biases (3-sample test for equality of proportions, $P_{adj}>0.05$ for each threshold distance; fig. 2b and supplementary table 4, Supplementary Material online), partially resulting from any formation of additional clusters being offset by the merging of adjacent clusters existing at lower threshold distances. Nevertheless, at all threshold distances, there are significant differences in the proportion of clusters across expression biases, with larva-preferentially expressed genes forming the largest number of clusters (4-sample test for equality of proportions, $P_{adj}<1\text{x}10^{-3}$ for each threshold distance; fig. 2b and supplementary table 5, Supplementary Material online).

Overall, the dichotomy shown by developmental stage-regulated genes in relation to the size and number of clusters is reflected in the fact that the quasi-monotonically decreasing trend in number of clusters as cluster size increases is not precisely mirrored by the number of clusters when these harbor ≥10 genes at any of the three threshold distances considered (fig. 3).

Almost half of the monarch chromosomes are enriched for life-stage and sex-specific expression biases

Under the null hypothesis, no evidence of enrichment for developmental stage-regulated genes should be found at a whole-chromosome scale. Our approach should nevertheless detect the known enrichment for male-biased genes on the ancestral (*anc*) but not on the novel (*neo*)

179 portion of the *Z* of *D. plexippus* (Ranz, et al. 2021b), thus reflecting incomplete dosage
180 compensation in the former but not in the latter portion of this heterochromosome (Gu, et al.
181 2019; Ranz, et al. 2021b). In Lepidoptera, females are the heterogametic sex so that genes on
182 the anc*Z* are expected to be underexpressed in female versus male tissues in the absence of
183 complete dosage compensation (Mank 2013).

184 We performed Monte Carlo simulations in which the gene order was shuffled across
185 chromosomes, finding 15 and 20 instances of significant enrichment and depletion at whole-
186 chromosome scale, respectively, for particular developmental stage-regulated gene categories
187 (fig. 4a and supplementary table 6, Supplementary Material online). In total, 21 out of 31
188 chromosomal elements showed patterns of nonrandom gene distribution, with 13 of them
189 (~42%) showing patterns of enrichment for particular expression biases. Eleven chromosomes
190 showed exclusivity in the kind of overrepresentation, *e.g.* only enrichment for larva-biased
191 expressed genes (chromosome *28*), as opposed to a mix of patterns of enrichment such as in
192 the case of chromosome *11*, which is enriched for both larva- and pupa-biased expressed
193 genes.

194 Among the patterns of enrichment was the expected excess of adult male- and deficit of adult
195 female-biased genes on the *anc-Z* (fig. 5). Unexpectedly, however, we also found enrichment
196 for male-biased genes in autosomes *24*, *26*, and *30*, and enrichment for female-biased genes in
197 autosomes *2*, *21*, and *27*. Departures from the random expectation were substantial in many
198 cases. For example, in the case of enrichment for male-biased genes, chromosome *24, 26,* and
199 *30* harbored 122%, 194%, and 282%, respectively, more of such genes than expected by
200 chance (observed vs average expected in the simulation data: 34 vs 15; 30 vs 10; and 43 vs 11,
201 respectively). For these autosomes, and unlike with the anc*Z*, incomplete dosage compensation
202 cannot be invoked to explain male-biased expression.

The nonrandom patterns documented are robust to the exclusion of lncRNAs, as shown by an additional set of simulations (fig. 4a and supplementary table 6, Supplementary Material online). Equally important, our ability to detect deviating patterns in relation to the random expectation was not impacted by the unequal number of genes across the four categories considered as shown by the lack of correlation between such number and the number of deviating cases, both when depletion or enrichment at whole-chromosome scale are considered jointly or separately (fig. 4b). Together, these findings are suggestive of a preferential localization of developmental stage-regulated genes with given expression biases in almost half of the *D. plexippus* chromosomes.

Enriched chromosomes for developmental stage-regulated genes tend to harbor larger than expected gene clusters

Whole-chromosome enrichment for developmental-stage regulated genes might be associated with gene clustering at a fine scale as the probability of physical aggregation should increase with the number of these genes on the same chromosome. Accordingly, we investigated whether the relationship between the overrepresentation of these genes at whole-chromosome scale translated into a proclivity to form more clusters, larger clusters, or both, in relation to the random expectation, this time paying attention to each individual chromosome. For the 15 cases of whole-chromosome scale enrichment for developmental stage-regulated genes, and the three threshold distances previously indicated, Monte Carlo simulations in which the gene order was shuffled within each chromosome showed that the observed number of clusters never exceeded the random expectation (P_{adj}>0.05 across individual chromosomes and threshold distances; supplementary table 7, Supplementary Material online). Reassuringly, none of the remaining whole-chromosome by expression bias combinations, *i.e.* those not showing evidence of significant global enrichment, showed a higher number of clusters than expected by chance.

But beyond the number of clusters, other metrics including the number of genes in clusters and the average cluster size might differ between whole-chromosome by expression bias combinations showing enrichment for developmental stage-regulated genes and those combinations not showing enrichment. We tested this possibility, finding a higher fraction of genes forming clusters in chromosomes showing such enrichments than expected, although only at a threshold distance ≤1 (P_{adj}=0.011; P_{adj}>0.05 for other threshold distances; supplementary table 8, Supplementary Material online). The only consistent difference across threshold distances between chromosomes enriched for particular categories of developmental stage-regulated genes and those that are not enriched was that the former harbored significantly larger clusters than the observed average size (P_{adj}<0.05 across distances; supplementary table 8, Supplementary Material online). Collectively, these results suggest that the above reported tendency of particular categories of developmental stage-regulated genes to form larger clusters than expected by chance at a fine-chromosome scale is tightly associated with the overrepresentation of such genes at a whole-chromosome scale.

Paralogous genes contribute to nonrandom patterns of fine-scale gene clustering but not to those at the whole-chromosome scale

Duplications bursts can contribute to the formation of large clusters of genes (Laukaitis, et al. 2008; Shipilina, et al. 2022), which can display a similar expression bias during particular life-stages if they retain common *cis*-regulatory sequences (Boutanaev, et al. 2002; Kustatscher, et al. 2017). We assessed how the presence of paralogs contributed to the nonrandom patterns of gene organization found at fine- and whole-chromosome scales. Here, paralogs are those that belong to the same orthogroup as delineated previously by OrthoFinder (Ranz, et al. 2021b). We found at least two or more paralogs in 32.6% (157/481) clusters at a distance of ≤1 intervening gene, a percentage that decreases slightly but significantly at higher distances (26.1% at ≤5 and 26.0% at ≤10; 3-sample test for equality of proportions, χ^2=7.38, d.f.=2,

253 P=0.025), although not when examined per each expression bias separately (P_{adj}>0.05 for each
254 threshold distance; supplementary fig. 1 and table 9, Supplementary Material online). We
255 noticed, nevertheless, significant differences in the relative presence of paralogs across clusters
256 with different expression biases although only at a threshold distance of ≤10 intervening genes
257 (Chi-square test of independence with simulated p-value based on 2000 replicates; distance ≤1:
258 χ^2=7.18, P_{adj}=0.054; distance ≤5: χ^2=8.12, P_{adj}=0.054; distance ≤10: χ^2=12.18, P_{adj}=0.022). At a
259 threshold distance ≤10, subsequent post-hoc tests indicated that paralogs are overrepresented
260 in clusters of genes preferentially expressed during the larval stage (P_{adj}=9.0×10^{-3}), and
261 underrepresented in clusters harboring genes preferentially expressed in adult males
262 (P_{adj}=0.039), relative to paralogs present in clusters of genes preferentially expressed during the
263 pupa stage or in adult females.

264 But the absence of reproducible discordances in the representation of paralogs in clusters of
265 different categories of developmental stage-regulated genes across threshold distances does
266 not directly address whether paralogs from the same orthogroup and with the same expression
267 bias are present in a higher number of clusters, or in more proportion in such clusters, than
268 expected by chance. To test these possibilities, we performed new sets of Monte Carlo
269 simulations in which genes were shuffled within chromosomes, and then the resulting gene
270 clusters, at the three distances, were inspected for the presence of paralogs from the same
271 orthogroup. We found that paralogs are part of more clusters, and represent a larger fraction of
272 genes in such clusters, than expected by chance. For example, at a distance ≤1, and
273 considering all expression biases jointly, 7-8 clusters should harbor essentially 15 paralogs
274 whereas the actual number of clusters with paralogs is 157, which include 446 paralogs. These
275 deviations from the random expectation are reproducible across threshold distances and when
276 the expression biases are considered separately (P_{adj}<1.0×10^{-5} for both metrics; supplementary
277 table 10, Supplementary Material online).

Paralogous genes with a similar expression bias during development also have the potential to contribute to the whole-chromosome scale enrichments documented. To assess this possibility, we performed additional Monte Carlo simulations in which redundant paralogous genes, *i.e.* those on the same chromosome and identical expression bias, were omitted. Two thirds of the original cases of whole-scale chromosome enrichment, 10 in total, were still significant (fig. 4a; supplementary table 6, Supplementary Material online). This includes all four instances of enrichment for male-biased genes and three of the five cases of enrichment for pupa-biased genes in expression. These results stress that the presence of unrelated genes with identical developmental-stage regulation on the same chromosome is a more relevant factor explaining the whole-chromosome enrichment patterns documented than the presence of paralogous genes with identical expression bias.

Common chromosome features fail to explain whole-scale chromosome enrichment for particular classes of developmental stage-regulated genes

We performed a multiple logistic regression analysis to examine more generally the impact of some basic features of the *D. plexippus* chromosomes and their evolutionary dynamics on the patterns of whole-chromosome scale enrichment for developmental stage-regulated genes. We considered chromosome size measured in Mb, chromosome size measured as number of genes, the number of preferentially expressed genes, the joint proportion of all types of developmentally biased genes in expression in relation to the total number of genes per chromosome, the fraction of repetitive sequences per chromosome, and the number of orthologs identified between *D. plexippus* and *B. mori*. We also included two measures of the rate of chromosomal rearrangement: the mere count of disruptions in gene order between the mentioned species as a proxy for the number of breakpoints of chromosomal inversions, and a second estimate of this number using a maximum parsimonious approach (Ranz, et al. 2022). And for the sake of completeness, we also considered the share of developmental stage-

regulated paralogs with the same expression bias that form clusters, as well as the proportion of paralogs regardless of their specific expression bias relative to all developmental stage-regulated genes on the same chromosome as additional predictors. As expected by the results above, the metrics related to paralogous genes failed to predict significantly whole-chromosome enrichment patterns. And among the rest of predictors, only the proportion of developmental stage-regulated genes in expression per chromosome was found to be significant (P=9.8×10^{-4}, supplementary table 11, Supplementary Material online). Chromosomes enriched for developmental stage-regulated genes harbor a 15% higher median proportion of these genes relative to non-enriched chromosomes (supplementary fig. 2, Supplementary Material online).

Large gene expression clusters often include paralogs of different multigene families

To better understand the functional and compositional characteristics of clusters of developmental stage-regulated genes, we targeted clusters including at least one more gene than the maximum threshold distance, *i.e.* 11 genes. At the threshold distances of 1, 5, and 10 intervening genes, we found 6, 10, and 15 of such clusters, respectively. To maximize the possibility of identifying robust patterns of shared properties among constituent genes, we focused on the clusters delineated at the highest threshold distance (table 1).

Close examination of the focal clusters clearly substantiated that a fraction of the constituent genes within clusters shared functional attributes such as their molecular function (supplementary table 12, Supplementary Material online). This observation is partially explained by the fact that these clusters are often populated by paralogs, as it happens for example with a cluster of 11 genes preferentially expressed in females that resides on chromosome *2*. This cluster spans 68.8 kb and BLASTP homology searches against *B. mori* and *D. melanogaster* revealed that eight of these genes encode cysteine proteinases (GO:0004197), all being part of the same orthogroup (Ranz, et al. 2021b). In other cases, the clusters include a mix of unrelated genes plus paralogs of different orthogroups. For example, chromosome *11* harbors a cluster of

32 genes preferentially expressed in pupa, spanning 293.6 kb. BLASTP homology searches revealed that many of these genes encode cuticular related proteins (GO:0040003). This cluster accommodates paralogs from two different orthogroups (12 and 4 respectively) among other unicopy genes. Likewise, chromosome *22* harbors a cluster of 18 genes, spanning 530.7 kb. This cluster includes many members of the ancient *Osiris* multigene family, which encodes plasma membrane proteins relevant for immunity and development (Smith, et al. 2018), and whose structural integrity has been documented in several insect lineages (Shah, et al. 2012). Some of the constituent genes are assigned to different orthogroups.

Further, we evaluated whether the clusters considered were located on evolutionary stable chromosomal regions within the Lepidoptera, *i.e.* those not impacted by chromosomal breakpoints fixed between *D. plexippus* and *B. mori* (Ranz, et al. 2022). If not located on stable regions, a fraction of the linked orthologs in *D. plexippus* should map in different microsynteny blocks, denoting that they are the byproduct of chromosomal rearrangements, possibly inversions. Eight of the 15 clusters harbor enough genes with 1-to-1 orthologs in *B. mori* as for being evaluated (Ranz, et al. 2022). For five of them, their constituent genes map on the same microsynteny block, one more cluster accommodates just one breakpoint in a terminal location, and the other two harbor multiple breakpoints (table 1). This accommodation of breakpoints does not seem correlated with a larger cluster size measured in kb (Pearson's r=0.585, P=0.128). For example, the cluster on chromosome *30* (table 1), the fourth smallest one among the eight clusters, accommodates multiple breakpoints. Our results suggest that at least a fraction of the largest coexpression clusters in *D. plexippus* reside in structurally dynamic genomic regions.

DISCUSSION

By exploiting the recently generated 'omic resources in *D. plexippus,* we investigated nonrandom gene distribution patterns at fine and whole-chromosome scales. At a fine-scale, and at three considered threshold distances between neighboring developmental stage-regulated genes, we document significant clustering, as previously reported in *Drosophila* species (Boutanaev, et al. 2002; Mezey, et al. 2008; Spellman and Rubin 2002; Weber and Hurst 2011) and *A. mellifera* (Duncan, et al. 2020), suggesting a common property of holometabolous insect genomes. In such clusters, roughly one third of the constituent genes are paralogs. This disproportionate number of paralogs in clusters of developmental stage-regulated genes is compatible with such clusters being unaffected by chromosomal rearrangements. In some cases, this could be just the result of lack of occurrence of structural rearrangements (Negre and Ruiz 2007), while in others the paralogs would remain nearby because of the detrimental effects of separating them if some sort of coordinated regulation exists among some of the paralogs. Further, most of the largest clusters harbor paralogs from not one but several orthogroups, as well as unrelated unicopy genes, which is suggestive of their accrual via chromosome rearrangements (Wong and Wolfe 2005), a pattern also documented for particular types of coexpression clusters in *A. mellifera* (Duncan, et al. 2020). Equally important, the constituent genes of some clusters map onto different microsynteny blocks between *D. plexippus* and *B. mori*, reminiscent of previous observations among *Drosophila* species (Weber and Hurst 2011). This begs the question about what fraction of the similarity in expression trend among the constituent genes of these clusters reflects *bona fide* functional coregulation as opposed to incidental coexpression as a result of being part of the same chromatin domain or preserving identical *cis*-regulatory sequences (Kustatscher, et al. 2017; Meadows, et al. 2010).

Unexpectedly, we uncovered a conspicuous preferential localization of developmental stage-regulated genes in roughly half of the *D. plexippus* chromosomes. These chromosomes, as usual in many Lepidoptera, are much smaller in size compared to those of *Drosophila* species

and *A. mellifera*, raising the opportunity of some degree of chromosome specialization. Notably, the relative presence of paralogs in those chromosomes and their tendency of such paralogs to cluster failed to explain overall enrichment at this chromosome scale. In different species, asymmetrical distributions of sex-biased genes in expression between the *X* and the *Z* heterochromosomes in relation to the autosomes have been reported. In *D. melanogaster*, it is known the under and overrepresentation of male-biased on the *X* and 2L autosome, respectively (Assis, et al. 2012; Meisel, et al. 2012; Ranz, et al. 2003). In *D. plexippus,* it is also known the enrichment for male- and female-biased genes on the anc-*Z* and autosomes as whole, respectively (Ranz, et al. 2021b). The patterns can be explained by a combination of factors with varying relevance across lineages: incomplete or absent dosage compensation; heterochromosome inactivation at the onset of meiosis; and sexually antagonistic alleles (Mahadevaraju, et al. 2021; Rice 1984; Vicoso and Charlesworth 2009; Wu and Xu 2003). Here, by analyzing each chromosome separately, three autosomes, and not only the *anc*-Z, were found significantly enriched for male-biased genes while another three autosomes were enriched for female-biased genes. These patterns could partially result from sexually antagonistic mutations with different degrees of dominance (h>0.5, for those female-beneficial; h<0.5, for those male-beneficial) becoming preferentially fixed on particular chromosomes. This possibility is not mutually exclusive from other mechanisms such as copy number increase of sex-biased genes via duplication in particular chromosomes. When comparing these two gene classes for the mentioned autosomes, duplication events seem to have a more substantial impact on whole-chromosome scale enrichment for female-biased than male-biased genes (15 out of 29 female-biased genes vs 8 out of 102 male-biased genes; 2-tailed Fisher's exact test, $P=7.41\times10^{-7}$).

But the chromosome enrichment patterns found here transcend sex-biased gene expression. Therefore, what mechanisms might be responsible for the preferential localization of

developmental stage-regulated genes on *D. plexippus* chromosomes? Synteny conservation in the Lepidoptera could be a contributing factor (Ahola, et al. 2014; Heliconius Genome, et al. 2012; Hill, et al. 2019; Ranz, et al. 2022). In the butterfly *Pieris napi*, a species in which synteny conservation has been eroded by interchromosomal rearrangements, conserved collinear blocks of genes show that such genes are enriched for particular Gene Ontology terms (Hill, et al. 2019). Similarly, the preservation of an initial gene content in particular chromosomes of *D. plexippus* could have enabled the subsequent accumulation of regulatory mutations that shaped the expression profiles of some of the resident genes. This would have facilitated the establishment of regulatory dependencies, with some taking place over long physical distances (Ghavi-Helm, et al. 2014; Kustatscher, et al. 2017). This process would have reinforced even further synteny conservation. Taken together, previous observations in *P. napi* (Hill, et al. 2019), and those here in *D. plexippus,* suggest the existence of nonrandom patterns of gene organization associated with gene functionality at least in some Lepidoptera.

Whether a mere consequence of synteny conservation or as reinforcing mechanism, the findings in *D. plexippus* warrants further examination across the Lepidoptera, which will require comprehensive transcriptome atlases in additional species. Finding pervasive phylogenetic evidence of an identical preferential chromosomal localization for particular developmental stage-regulated genes would argue for the importance of the interplay between gene expression profiles and chromosome organization across Lepidoptera. Failing to find such evidence would be indicative of a more phylogenetic restricted pattern. Further, similar analyses to ours will have to be extended to genes preferentially expressed during embryogenesis as these genes were not part of the transcriptome atlas generated (Ranz, et al. 2021b). Our results not only enhance our understanding on the relationship between gene expression bias during development and chromosome function, but open new avenues of inquiry about what factors might be influencing Lepidoptera chromosome organization.

METHODS

Gene information. Gene information about preferential expression at a particular broadly defined developmental stage in *D. plexippus* was previously determined based on RNA-seq data used to generate a transcriptome atlas in this species (Ranz, et al. 2021b). Here, we used the lists of preferentially expressed genes delineated in three of the contrasts performed: *Lpool:Other, Ppool:Other,* and *Sexes* (supplementary table 7 in Ranz, et al. 2021b). These contrasts compared expression levels from RNA-seq experiments derived from pools of larvae from three different stages (*Lpool:* L1, L3, L5), pools of pupae from five different stages (*Ppool:* P1, P3, P5, P7, P9), and adult male and female (*Mpool* and *Fpool,* respectively) anatomical parts (heads, thorax, and abdomen). For each pool, aliquots from the different contributing samples were mixed equimolarly. In the contrast *Lpool:Other*, the expression level across larva stages was tested for differences in relation to the expression across pupa stages and adult parts. The same logic follows for the contrast *Ppool:Other*. From these two contrasts, we retrieved the list of genes preferentially expressed during the larva and pupa stages, respectively. In the contrast *Sexes,* the expression level across anatomical parts of adult individuals was compared between males and females. As some genes could be for example not only pupa preferentially expressed but also male preferentially expressed, we omitted them in downstream analysis. Therefore, we only included in our analyses those genes preferentially expressed during the larva and pupa stages but not sex-biased during adulthood (1178 and 891, respectively), plus those either male- or female-biased during adulthood but not peaking in expression during the larva or pupa stages relative to adulthood (582 and 155, respectively). From these, 2732 genes were found in scaffolds reliably mapped to the chromosomes of *D. plexippus*, thus excluding 4, 56, and 14 genes preferentially expressed during the larva, pupa, and adult male stages that could not be reliably mapped onto particular chromosomes (Ranz, et al. 2021b). In the three contrasts, differential expression between the two conditions under

comparison was determined with gmlTreat (McCarthy and Smyth 2009) within the edgeR package (McCarthy, et al. 2012), thus requesting both a differential expression higher than a log2(fold-change) of |1| and a 5% FDR (Ranz, et al. 2021b). As a result, the genes found to be differentially expressed show typically log(fold-changes) higher than the threshold selected. The reproducibility and reliability of the expression data associated with the three contrasts used here is confirmed by the strong correlation between biological replicates within the same sample type (Pearson's *r*=0.96, 0.947, 0.981, 0.972 for *Lpool*, *Ppool*; *Mpool*, and *Fpool* samples, respectively) and the correct clustering of biological replicates in relation to developmental stage in the multidimensional scaling analysis performed (supplementary fig. 16b in Ranz, et al. 2001b).

Genes related by duplication events, *i.e.* those part of the same orthogroup according to OrthoFinder (Emms and Kelly 2015), were as delineated (Ranz, et al. 2021b), and are available through Zenodo (Gonzalez-De-la-Rosa, et al. 2021). Disruption of microsynteny as a result of chromosomal rearrangements occurred during the evolution of the lineages that lead to *D. plexippus* and *B. mori* relied on chromosome positional information previously generated (Ranz, et al. 2021b), and provided through Dryad (Ranz, et al. 2021a).

Functional characterization of genes in clusters was done by performing BLASTP searches against *B. mori, D. plexippus,* and *D. melanogaster* in Ensembl Metazoa (Cunningham, et al. 2022), as of Mar 10th 2023, and integrating the existing functional annotations from the best BLASTP hits documented, always with E-values lower than 1.0E-05, across the three species.

Tests of nonrandom chromosomal distribution of developmental stage-regulated genes. To determine whether genes preferentially expressed in a given developmental stage are distributed randomly across the chromosomes of *D. plexippus,* we performed 1×10^5 permutations. In each permutation, we shuffled without replacement the actual chromosomal location of 14,207 coding and 478 lncRNA annotated genes, whether preferentially expressed

or not (Ranz, et al. 2021b). The number of genes per chromosome was kept as in the actual data across permutations. From the permutations performed, we generated the expected null distributions for the gene counts of each chromosome and expression bias combination. Then, for each null distribution, we recorded in how many permutations the number of genes with a given expression bias was the same or higher (or the same or lower) than that observed. The fraction of these permutations in relation to the total number of replicates was taken as the exact probability of enrichment (or depletion) of genes with a given expression bias in each chromosome. Subsequently, the list of 248 (31 chromosomes $\times$ 4 expression biases $\times$ 2 directions) *p*-values were corrected for multiple comparisons (Benjamini and Hochberg 1995). A 5% FDR was applied to call for whole-chromosome enrichment (or depletion) of genes with a particular expression bias. Two additional sets of similar simulations were performed by omitting particular genes: in one, all lncRNAs; and in the other, all redundant paralogs on the same chromosome, *i.e.* those showing the same expression bias, but one, which was chosen at random.

To test for nonrandom patterns of gene organization at a fine chromosome scale, 1×10^5 permutations were performed in which gene shuffling was done within each chromosome. In each of these permutations, we recorded the number of clusters of developmental stage-regulated genes and the number of genes in such clusters, allowing the estimation of the average cluster size and the size of the largest cluster. This was done at three threshold distances between adjacent genes in clusters, in which the distance is measured as the maximum number of intervening genes (1, 5, and 10). For the indicated parameters, threshold distance, and chromosome by expression bias combination, we generated the expected null distributions, and following the above rationale for deviating patterns at whole-chromosome scale, we determined the probability of obtaining values equal or higher than those observed at a 5% FDR. To calibrate the contribution of paralogous genes on nonrandom patterns of gene

organization at a fine scale, another set of simulations was performed in which the average number of clusters with developmental stage-regulated paralogs and the average number of such genes within clusters were examined.

Statistical analysis. Permutation simulations, regression analysis, and chi-square-based test for inequality of proportions were performed using built-in functions in R (R Development Core Team 2016).

ACKNOWLEDGEMENTS

We thank John Parsch, Andrey Przhibelskiy, Stephen Schaeffer, and Igor Sharakhov for comments on the manuscript and technical aspects related to our analyses, and Jaylen Lee, from the Biostatistics, Epidemiology, and Research Design Unit at UC Irvine ICTS, and Laurence Mueller for feedback on simulation analyses. This work was supported by a UC Mexus and an National Sciences Foundation (MCB2129845) awards (J.M.R).

DATA AVAILABILITY

No new data were generated by this research. Scripts and input file for the different types of data permutation analyses performed are available at Zenodo (Kimura and Ranz 2023) (10.5281/zenodo.10059296).

AUTHOR CONTRIBUTIONS

J.M.R. conceived, designed, and supervised the experiments as well as wrote the manuscript. T.A.M. supervised and validated the experiments and wrote the manuscript. A.K., A.C.G., and J.M.R. performed the analyses.

COMPETING INTERESTS

The authors declare no competing interests.

Table 1. Salient features of the largest clusters of developmental stage-regulated genes in *D. plexippus*

D. plexippus **Chromosome**	**Expression Bias**	**Genes**	**Artifact Risk ***	**Size (kb)**	***Bmo*** **Orthologs Identified †**	**Microsynteny Blocks ‡**	**Orthogroup Composition §**	**Conservation ¦**	**Functional Signatures ‖**
2	F	11	No	68.83	0	Na	8	Not evaluable	Cysteine-type endopeptidase activity
3	L	16	No	190.718	1	Na	15	Not evaluable	Not evaluable
11	L	13	No	279.99	4	1	8	Yes	Chitin-based cuticle development, GATA transcript Rho activating protein
11	L	21	No	620.73	5	1	6	Yes	Chitin-based cuticle development
11	P	32	No	293.61	8	1	4, 12	Yes	Chitin-based cuticle development
12	P	32	No	316.41	6	1	2, 5, 5	Yes	Cuticle protein
14	P	17	No	778.25	4	2	2, 3	Almost complete	Hormone binding/protein takeout (circadian clock-controlled), cuticle protein
16	L	14	No	236.009	1	Na	0	Not evaluable	Metamorphosis
17	L	12	Yes	Uncertain	1	Na	2, 4	Not evaluable	Protein metabolism, membrane trafficking
19	L	14	Yes	Uncertain	4	3	2, 2	Not evaluable	Cytochrome P450
22	P	18	No	530.65	10	1	2	Yes	Transmembrane proteins
26	M	18	No	1007.1	3	3	0	No	Not evaluable
28	L	25	Yes	Uncertain	0	Na	2, 2, 4, 5, 6	Not evaluable	Carbohydrate metabolism, histones
28	L	18	Yes	Uncertain	0	Na	3, 2, 9	Not evaluable	Histones
30	L	12	No	418.52	3	3	3	No	Nuclear pore complex protein Nup98-Nup96, protein metabolism, cuticle protein

Bmo, B. mori. L, P, M, and F refer to the overexpression bias documented (larva, pupa, adult male, adult female, respectively).

* Artifact risk is associated with constituent genes mapping into two joined but different contigs (Ranz, et al. 2021b). As a result, the size of the cluster could not be reliably estimated.

† For the constituent genes in the cluster of *D. plexippus* as reported (Ranz, et al. 2021b).

‡ Number of microsynteny blocks where the genes of *D. plexippus* with orthologs in *B. mori* are located (Ranz, et al. 2022). Na, not analyzable because 0-1 orthologs were detected in *B. mori.*

§ Number of duplicates in orthogroups represented with >1 gene. The number of duplicates in different orthogroups are separated by commas. 0, no two constituent genes were identified as part of the same orthogroup (Ranz, et al. 2021b).

¦ Not evaluable either because of <2 orthologs were detected in *B. mori*, the possible artifactual nature of the cluster, or both.

‖ Recurrent functional annotation among the genes present in the cluster. Not all the genes in the cluster necessarily have the same functional properties. Not evaluable, absence of functional information or nonsignificant BLASTP hits for the constituent genes. For detailed gene-by-gene information, see supplementary table 12, Supplementary Material online.

FIGURE LEGENDS

Figure 1. Chromosomal distribution of genes with preferential expression in one of the broadly defined developmental life stages considered in *D. plexippus*.

Genes with different expression biases are color coded (L, larva; P, pupa; M, adult male; F, adult female). The graph was generated with the online tool PhenoGram (http://visualization.ritchielab.org/).

Figure 2. Developmental-stage regulated genes form clusters across the chromosomes of *D. plexippus.*

a Percentage of developmental stage-regulated genes that are part of clusters across three threshold distances. The distance between constituent genes of clusters is measured as the number of intervening genes between them, which can be 1 or less, 5 or less, or 10 or less. Significant differences were found in the proportion of clustered genes across developmental expression biases for each threshold distance (supplementary table 2, Supplementary Material online), and in the proportion of clustered genes across threshold distances for each expression trend but for female-preferentially expressed genes (supplementary table 3, Supplementary Material online). **b** Percentage of clusters harboring developmental stage-regulated genes with a given expression bias relative to the total number of clusters at each threshold distance, which is indicated on top. The proportion of clusters does not increase significantly with the threshold distance for any of the expression biases (supplementary table 4, Supplementary Material online), whereas, for each threshold distance, there are significant differences in the proportion of clusters among expression biases (supplementary table 5, Supplementary Material online). The different developmental expression biases are color coded (top legend): L, larva; P, pupa; M, adult male; F, adult female.

Figure 3. Magnitude of the clustering of developmental-stage regulated genes within the chromosomes of *D. plexippus*.

a Number of genes in clusters. **b** Number of clusters. Both metrics are plotted as a function of cluster size, which is measured in number of constituent genes. These clusters were delineated at three different threshold distances (from top to bottom) between adjacent genes part of such clusters. This distance was measured as the number of intervening genes: 1 or less, 5 or less, 10 or less. Clusters harboring 10 or more genes were grouped. The total number of genes clustered and the number of clusters at each threshold distance are indicated on the top center of each chart. The different developmental expression biases are color coded (top right legend): L, larva; P, pupa; M, adult male; F, adult female.

Figure 4. Number of cases of nonrandom organization of developmental-stage regulated genes at whole-chromosome scale in *D. plexippus.*

a Number of chromosomes showing either depletion or enrichment for genes with preferential expression in larva (L), pupa (P), adult male (M); and adult female (F). Deviation from the random expectation was determined by performing Monte Carlos simulations (see supplementary table 6, Supplementary Material online, for details about specific chromosomes). The deviating patterns found when all genes are considered (left) are largely robust to the omission of lncRNAs (middle) and redundant paralogs, *i.e.* those from the same orthogroup, with the same expression bias, and on the same chromosome. **b** Linear regression analysis

between the number of genes in different categories of preferentially expressed genes and the number of deviating cases from the random expectation. The results for enrichment only, depletions only, and the combination of both patterns are shown along with their corresponding coefficients of determination and statistical significance. In no case, is the number of genes showing a particular expression bias correlated with the number of cases in which chromosomes show deviating patterns (*i.e.* enrichment or depletion) in relation to the random expectation.

Figure 5. Null distributions and actual number of resident genes on the anc-*Z* chromosome showing particular trends of preferential expression during the life cycle of *D. plexippus*.

The null distributions for the number of genes showing different expression biases (L, larva; P, pupa; M, adult male; F, adult female) according to Monte Carlo simulations (n=100,000) are color coded as in fig. 1. Where, within the null distribution, the observed number of genes with a given expression bias falls is indicated above the arrowhead. The adjusted probability of finding the observed value or lower in the case of underrepresentation, or higher in the case of overrepresentation, relative to the random expectation at whole-chromosome scale are shown. The number of genes found in simulations 100 and 99,900 upon sorting the values from lower to higher are provided below the null distributions.

REFERENCES

Ahola V, et al. 2014. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. Nat Commun 5: 4737. doi: 10.1038/ncomms5737

Assis R, Zhou Q, Bachtrog D 2012. Sex-biased transcriptome evolution in Drosophila. Genome Biol Evol 4: 1189-1200. doi: 10.1093/gbe/evs093

Catalogue of Life Checklist [Internet]. 2022. Available from https://www.catalogueoflife.org/.

Benjamini Y, Hochberg Y 1995. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 57: 289-300.

Bonasio R, Shiekhattar R 2014. Regulation of transcription by long noncoding RNAs. Annu Rev Genet 48: 433-455. doi: 10.1146/annurev-genet-120213-092323

Boutanaev AM, Kalmykova AI, Shevelyov YY, Nurminsky DI 2002. Large clusters of co-expressed genes in the *Drosophila* genome. Nature 420: 666-669. doi: 10.1038/nature01216

Cunningham F, et al. 2022. Ensembl 2022. Nucleic Acids Res 50: D988-D995. doi: 10.1093/nar/gkab1049

de Vos JM, Augustijnen H, Batscher L, Lucek K 2020. Speciation through chromosomal fusion and fission in Lepidoptera. Philos Trans R Soc Lond B Biol Sci 375: 20190539. doi: 10.1098/rstb.2019.0539

Duncan EJ, Leask MP, Dearden PK 2020. Genome Architecture Facilitates Phenotypic Plasticity in the Honeybee (Apis mellifera). Mol Biol Evol 37: 1964-1978. doi: 10.1093/molbev/msaa057

Ellis EA, Storer CG, Kawahara AY 2021. De novo genome assemblies of butterflies. Gigascience 10. doi: 10.1093/gigascience/giab041

Emms DM, Kelly S 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol 16: 157. doi: 10.1186/s13059-015-0721-2

Ghavi-Helm Y, et al. 2014. Enhancer loops appear stable during development and are associated with paused polymerase. Nature 512: 96-100. doi: 10.1038/nature13417

Gonzalez-De-la-Rosa PM, Ranz JM, Abreu-Goodger C. 2021. Danaus plexippus genome annotation and orthofinder results. In. . Zenodo.

Gu L, et al. 2019. Dichotomy of Dosage Compensation along the Neo Z Chromosome of the Monarch Butterfly. Curr Biol 29: 4071-4077 e4073. doi: 10.1016/j.cub.2019.09.056

Heliconius Genome C, et al. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487: 94-98.

Hill J, et al. 2019. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. Sci Adv 5: eaau3648. doi: 10.1126/sciadv.aau3648

Hurst LD, Pal C, Lercher MJ 2004. The evolutionary dynamics of eukaryotic gene order. Nat Rev Genet 5: 299-310.

Kimura A, Ranz JM. 2023. Testing non-random patterns of gene organization in Dannaus plexipus. Version 1.0.0. Zenodo.

Kustatscher G, Grabowski P, Rappsilber J 2017. Pervasive coexpression of spatially proximal genes is buffered at the protein level. Mol Syst Biol 13: 937. doi: 10.15252/msb.20177548

Laukaitis CM, et al. 2008. Rapid bursts of androgen-binding protein (Abp) gene duplication occurred independently in diverse mammals. BMC Evol Biol 8: 46. doi: 10.1186/1471-2148-8-46

Lee JM, Sonnhammer EL 2003. Genomic gene clustering analysis of pathways in eukaryotes. Genome Res 13: 875-882. doi: 10.1101/gr.737703

Lercher MJ, Urrutia AO, Hurst LD 2002. Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet 31: 180-183.

Lukhtanov VA, et al. 2018. Versatility of multivalent orientation, inverted meiosis, and rescued fitness in holocentric chromosomal hybrids. Proc Natl Acad Sci U S A 115: E9610-E9619. doi: 10.1073/pnas.1802610115

Mackintosh A, et al. 2022. The genome sequence of the lesser marbled fritillary, Brenthis ino, and evidence for a segregating neo-Z chromosome. G3 (Bethesda) 12. doi: 10.1093/g3journal/jkac069

Mahadevaraju S, et al. 2021. Dynamic sex chromosome expression in Drosophila male germ cells. Nat Commun 12: 892. doi: 10.1038/s41467-021-20897-y

Mandrioli M, Manicardi GC 2020. Holocentric chromosomes. PLoS Genet 16: e1008918. doi: 10.1371/journal.pgen.1008918

Mank JE 2013. Sex chromosome dosage compensation: definitely not for everyone. Trends Genet 29: 677-683. doi: 10.1016/j.tig.2013.07.005

Marec F 1996. Synaptonemal complexes in insects. International journal of insect morphology and embryology 25: 205-233. doi: 10.1016/0020-7322(96)00009-8

McCarthy DJ, Chen Y, Smyth GK 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 40: 4288-4297. doi: 10.1093/nar/gks042

McCarthy DJ, Smyth GK 2009. Testing significance relative to a fold-change threshold is a TREAT. Bioinformatics 25: 765-771. doi: 10.1093/bioinformatics/btp053

Meadows LA, Chan YS, Roote J, Russell S 2010. Neighbourhood continuity is not required for correct testis gene expression in Drosophila. PLoS Biol 8: e1000552.

Meisel RP, Malone JH, Clark AG 2012. Faster-X evolution of gene expression in Drosophila. PLoS Genet 8: e1003013. doi: 10.1371/journal.pgen.1003013

Mezey JG, Nuzhdin SV, Ye F, Jones CD 2008. Coordinated evolution of co-expressed gene clusters in the Drosophila transcriptome. BMC Evol Biol 8: 2.

Mongue AJ, Nguyen P, Volenikova A, Walters JR 2017. Neo-sex Chromosomes in the Monarch Butterfly, Danaus plexippus. G3 (Bethesda) 7: 3281-3294. doi: 10.1534/g3.117.300187

Negre B, Ruiz A 2007. HOM-C evolution in Drosophila: is there a need for Hox gene clustering? Trends Genet 23: 55-59. doi: 10.1016/j.tig.2006.12.001

Quintero-Cadena P, Sternberg PW 2016. Enhancer Sharing Promotes Neighborhoods of Transcriptional Regulation Across Eukaryotes. G3 (Bethesda) 6: 4167-4174. doi: 10.1534/g3.116.036228

R Development Core Team. 2016. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL 2003. Sex-dependent gene expression and evolution of the Drosophila transcriptome. Science 300: 1742-1745. doi: 10.1126/science.1085881

Ranz JM, Gonzalez-De-la-Rosa PM, Abreu-Goodger C. 2021a. One-to-one orthologs between Danaus plexippus and other insect species. In. Dryad.

Ranz JM, et al. 2021b. A de novo transcriptional atlas in Danaus plexippus reveals variability in dosage compensation across tissues. Commun Biol 4: 791. doi: 10.1038/s42003-021-02335-3

Ranz JM, et al. 2022. Multiscale analysis of the randomization limits of the chromosomal gene organization between Lepidoptera and Diptera. Proc Biol Sci 289: 20212183. doi: 10.1098/rspb.2021.2183

Rice WR 1984. Sex chromosomes and the evolution of sexual dimorphism. Evolution 38: 735-742.

Robinson R. 1971. Lepidoptera Genetics. Oxford: Pergamon.

Semon M, Duret L 2006. Evolutionary origin and maintenance of coexpressed gene clusters in mammals. Mol Biol Evol 23: 1715-1723.

Shah N, Dorer DR, Moriyama EN, Christensen AC 2012. Evolution of a large, conserved, and syntenic gene family in insects. G3 (Bethesda) 2: 313-319. doi: 10.1534/g3.111.001412

Shipilina D, et al. 2022. Linkage mapping and genome annotation give novel insights into gene family expansions and regional recombination rate variation in the painted lady (Vanessa cardui) butterfly. Genomics 114: 110481. doi: 10.1016/j.ygeno.2022.110481

Smith CR, Morandin C, Noureddine M, Pant S 2018. Conserved roles of Osiris genes in insect development, polymorphism and protection. J Evol Biol 31: 516-529. doi: 10.1111/jeb.13238

Spellman PT, Rubin GM 2002. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. J Biol 1: 5.

Szabo Q, Bantignies F, Cavalli G 2019. Principles of genome folding into topologically associating domains. Sci Adv 5: eaaw1668. doi: 10.1126/sciadv.aaw1668

Vicoso B, Charlesworth B 2009. Effective population size and the faster-X effect: an extended model. Evolution 63: 2413-2426. doi: 10.1111/j.1558-5646.2009.00719.x

Weber CC, Hurst LD 2011. Support for multiple classes of local expression clusters in Drosophila melanogaster, but no evidence for gene order conservation. Genome Biol 12: R23. doi: 10.1186/gb-2011-12-3-r23

Wen K, et al. 2016. Critical roles of long noncoding RNAs in Drosophila spermatogenesis. Genome Res 26: 1233-1244. doi: 10.1101/gr.199547.115

Williams EJ, Bowles DJ 2004. Coexpression of neighboring genes in the genome of Arabidopsis thaliana. Genome Res 14: 1060-1067.

Wong S, Wolfe KH 2005. Birth of a metabolic gene cluster in yeast by adaptive gene relocation. Nat Genet 37: 777-782.

Wu CI, Xu EY 2003. Sexual antagonism and X inactivation--the SAXI hypothesis. Trends Genet 19: 243-247.

Yasukochi Y, Ashakumary LA, Baba K, Yoshido A, Sahara K 2006. A second-generation integrated map of the silkworm reveals synteny and conserved gene order between lepidopteran insects. Genetics 173: 1319-1328. doi: 10.1534/genetics.106.055541

Zhu B, Li L, Wei R, Liang P, Gao X 2021. Regulation of GSTu1-mediated insecticide resistance in Plutella xylostella by miRNA and lncRNA. PLoS Genet 17: e1009888. doi: 10.1371/journal.pgen.1009888

Zinani OQH, Keseroglu K, Ozbudak EM 2022. Regulatory mechanisms ensuring coordinated expression of functionally related genes. Trends Genet 38: 73-81. doi: 10.1016/j.tig.2021.07.008